Computerized Medical Imaging and Graphics 70 (2018) 101–110

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks

Fan Li [a], Manhua Liu [a,b,*], the Alzheimer's Disease Neuroimaging Initiative

[a] Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] Shanghai Engineering Research Center for Intelligent Diagnosis and Treatment Instrument, Shanghai Jiao Tong University, Shanghai, 200240, China

ARTICLE INFO

Article history:
Received 5 April 2018
Received in revised form 12 July 2018
Accepted 26 September 2018

Keywords:
Alzheimer's disease
Dense convolutional network
K-Means clustering
Patch features
Structural magnetic resonance image

ABSTRACT

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder with progressive impairment of memory and cognitive functions. Structural magnetic resonance images (MRI) play important role to evaluate the brain anatomical changes for AD Diagnosis. Machine learning technologies have been widely studied on MRI computation and analysis for quantitative evaluation and computer-aided-diagnosis of AD. Most existing methods extract the hand-craft features after image processing such as registration and segmentation, and then train a classifier to distinguish AD subjects from other groups. Motivated by the success of deep learning in image classification, this paper proposes a classification method based on multiple cluster dense convolutional neural networks (DenseNets) to learn the various local features of MR brain images, which are combined for AD classification. First, we partition the whole brain image into different local regions and extract a number of 3D patches from each region. Second, the patches from each region are grouped into different clusters with the K-Means clustering method. Third, we construct a DenseNet to learn the patch features for each cluster and the features learned from the discriminative clusters of each region are ensembled for classification. Finally, the classification results from different local regions are combined to enhance final image classification. The proposed method can gradually learn the MRI features from the local patches to global image level for the classification task. There are no rigid registration and segmentation required for preprocessing MRI images. Our method is evaluated using T1-weighted MRIs of 831 subjects including 199 AD patients, 403 mild cognitive impairment (MCI) and 229 normal control (NC) subjects from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results show that the proposed method achieves an accuracy of 89.5% and an AUC (area under the ROC curve) of 92.4% for AD vs. NC classification, and an accuracy of 73.8% and an AUC of 77.5% for MCI vs. NC classification, demonstrating the promising classification performances.

1. Introduction

Alzheimer's disease (AD) is an irreversible brain disorder with progressive impairment of the memory and cognitive functions. It is the most common case of dementia in the late life of humans. Mild cognitive impairment (MCI) is a transitional state from healthy to dementia and it is usually considered as a clinical precursor of AD. Currently, there are no effective cure for AD. But some treatments can be developed to delay its progression, especially if AD can be diagnosed at an early stage. Thus, its early diagnosis is important for patient care and treatment. But it is still a challenging problem for accurate and early diagnosis of AD/MCI in clinic. Magnetic resonance images (MRI) including structural magnetic resonance images (sMRI) and functional MRI (fMRI) are non-invasive and powerful imaging tools to help understand and evaluate the anatomical and functional neural changes related to AD (Herrup, 2011; Jr et al., 2011; Liu et al., 2014). In recent years, extensive efforts have been done to develop computer-aided system using the various machine learning methods to decode the disease states with MR images (Herrup, 2011; Jr et al., 2011; Zhang et al., 2011; Liu et al., 2013; Suk et al., 2015).

Since the raw MR brain image is too huge to be directly used for classification, it is necessary to preprocess the MR images and perform the feature extraction and classification for disease diagnosis. One of the most widely used methods is to partition the image into multiple anatomical regions, i.e., regions of interest (ROIs), through the warping of a labeled atlas, and the regional measurements such as volumes are computed as the features for AD

* Corresponding author at: Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai, 200240, China.
E-mail address: mhliu@sjtu.edu.cn (M. Liu).

https://doi.org/10.1016/j.compmedimag.2018.09.009
0895-6111/

classification (Herrup, 2011; Zhang et al., 2011; Suk et al., 2015). For feature selection, a discriminative multi-task method was proposed to select the most discriminative features from 93 ROIs for multi-modality classification of AD/MCI (Ye et al., 2016). Furthermore, a hierarchical feature and sample selection framework was proposed to gradually select informative features from 93 predefined ROIs and discard ambiguous samples for improving classifier learning (Le et al., 2017). A multi-kernel learned method was combined with marginal fisher analysis to simultaneously select a subset of the relevant brain ROIs and learn a dimensionality transformation (Cao et al., 2017). In addition to the ROI features, deep learning networks were recently used to extract the latent high-level features from measurements of ROIs for AD classification (Zhang et al., 2011; Suk et al., 2015). A stacked autoencoder was investigated to learn the latent high-level features from ROIs for improvement of classification performance (Suk et al., 2015). A novel diagnostic framework with stacked autoencoder was proposed to learn high-level features of ROIs and with a zero-masking strategy for data fusion of multiple image modalities (Liu et al., 2015aa and b). Although promising results of brain image analysis have been reported, there are still some limitations in the ROI based methods. First, the definition of ROIs requires the accumulation of long-term experience of researchers. Second, the segmentation of ROIs is also affected by the individual differences and subjective factors of scientific researches. Third, the morphological abnormalities caused by the brain disorders do not always occur in the pre-defined ROIs, and they may involve multiple ROIs or part of the extracted ROI, so the performance may not be stable.

Instead of partitioning the brain image into ROIs, a landmark-based feature extraction method was proposed for fast AD diagnosis without nonlinear registration and tissue segmentation (Zhang et al., 2016). A number of landmark points were detected based on shape constraint and the morphological features were extracted from the landmarks to train a linear SVM classifier for AD diagnosis. Furthermore, the landmark-based method was extended for analysis of longitudinal MR images (Zhang et al., 2017). The high-level statistical spatial and contextual longitudinal features were extracted from the landmarks to capture the spatial structural abnormalities and longitudinal variations, which were input to train a linear SVM classifier for AD diagnosis. The circular harmonic functions (CHFs) were investigated to extract the local features from the most involved areas of the disease: Hippocampus and Posterior Cingulate Cortex (PCC) in three brain projections and classify the brain images (Ben et al., 2015).

In recent years, the deep learning methods have been widely investigated to jointly learn the features from the images and class discrimination for image classification and computer vision (Simonyan and Zisserman, 2014; Zhu et al., 2017). They also achieved great success to learn the feature and identify the patterns for medical image analysis and computer aided disease diagnosis (Shen et al., 2017). Different from the traditional methods that extracts the handcrafted features with domain specific knowledge, deep learning can construct a deep neural network architecture to learn the hierarchical representations from the raw image data. Thus, the complex patterns can be identified with deep learning. Convolutional neural networks (CNNs) were investigated to learn the features of MR brain images for AD diagnosis (Adrien, 2015, Hosseini-Asl et al., 2016). A deep 3D convolutional neural network (3D-CNN) was built upon a 3D convolutional Autoencoders to capture the anatomical shape variations of the structural MRI scans to predict AD (Hosseini-Asl et al., 2016). This method can learn the features from the raw image data to capture AD biomarkers and adapt to different domain datasets. A deep learning classification algorithm was proposed for AD diagnosis using both structural and functional MRI (Adrien, 2015). In this method, the CNN model was built with one convolutional layer trained with sparse Autoencoder, which was explored to extract the imaging features for AD classification. The above methods can learn the features capturing AD biomarkers via convolutional network. But they require the convolutional filters pretrained on Autoencoder with carefully preprocessed data to extract features and then classify them for task-specific target. A landmark based deep feature learning (LDFL) framework was proposed for automatic diagnosis of AD using MRI (Liu et al., 2018). A number of discriminative anatomical landmarks were identified in a data-driven manner and a set of patches were extracted from the landmarks to build a deep CNN for automatical extraction of patch-based representation from MRI. A novel deep ensemble sparse regression network was proposed that combines the sparse regression and deep learning for diagnosis and prognosis of AD and MCI (Suk et al., 2017). By regarding the response values of the sparse regression models as target-level representations, a deep CNN was built for clinical decision making. A classification method was proposed by ensemble of multiple deep 3D convolutional neural networks (3D-CNNs) to learn the various features from local brain regions for AD classification, which can alleviate the problem of small number of training samples (Cheng and Wang, 2017).

Recently, DenseNet (Huang et al., 2016) was proposed as a new structure of deep convolutional neural network, which connects each layer to every other layers in a feed-forward fashion to capture and reuse the rich features of different layers and thus achieves better performance than other CNN. Motivated by the success of DenseNet in computer vison, this paper proposes a novel classification method based on combination of multiple cluster DenseNets for MR brain image classification and disease diagnosis. Instead of extracting the region of interests (ROIs) predefined by human experts, we uniformly partition the whole brain image into $3 \times 3 \times 3$ different regions and a number of 3D patches are sampled and extracted from each region. Then, K-means clustering is applied to group the patches from one region into different clusters and a deep DenseNet is trained for each cluster. The features learned by multiple cluster DenseNets are aggregated for the representation of local region. Finally, classification results of multiple local regions are combined to make the final classification. Compared to the existing methods, our proposed method has the following advantages: 1) it can alleviate the problem of small image set on training DenseNet. Usually training a DenseNet requires a large image set, which is not applicable for AD diagnosis. Instead of training a deep DenseNet with the whole brain image, we can build a DenseNet on each cluster with a number of local image patches sampled from image region for network training. 2) No tissue and ROI segmentations are required in image processing, which can simplify the diagnosis procedure and save the computation costs. 3) No rigid registration is required before feature extraction, which can reduce the computation costs. Clustering is used to group similar image patches into clusters, which can achieve the robustness of image variances.

The rest of this paper is organized as follows. In Section 2, we present the materials and the proposed method in details. In Section 3, we provide the experiments and results. A conclusion will be given in Section 4.

2. Proposed method

In this section, we will present the proposed classification framework in detail. Our proposed method makes no assumption on a specific neuroimaging modality. The T1-weighted MR brain images are widely available, non-invasive and often used as the first biomarker in AD diagnosis. Thus, they are used to test the proposed method in this work. For brain image analysis, one direct way is to build a deep DenseNet with the whole 3D image for feature learn-

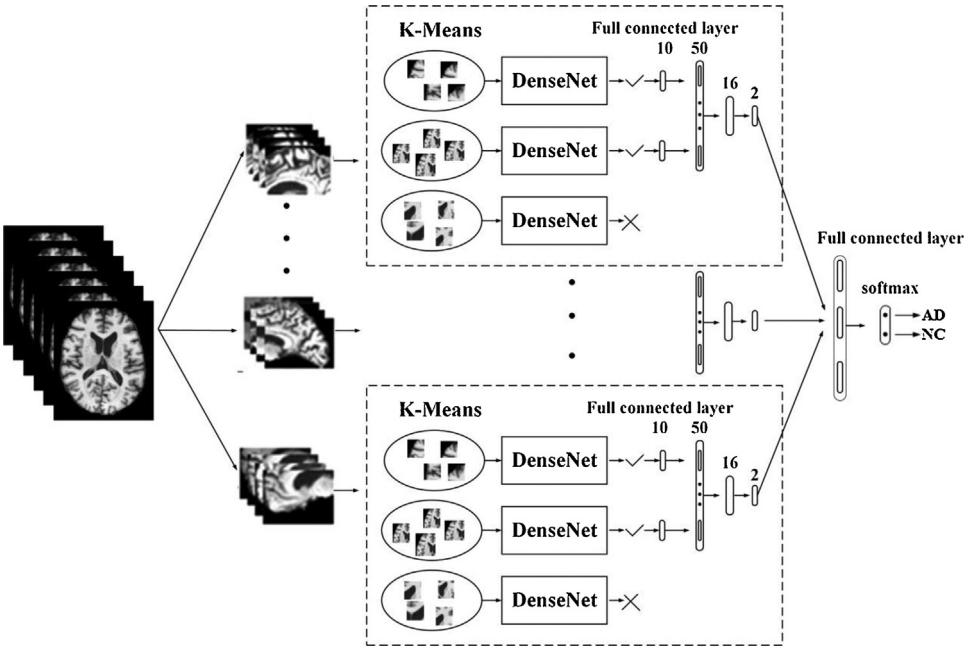

Fig. 1. An overview of our proposed classification framework which consists of image processing, patch extraction and clustering, construction of DenseNet for each cluster and final ensemble classification.

ing and classification jointly. However, training deep DenseNet may be challenged by the risk of over-fitting as the MR image sets for AD diagnosis are relatively small compared to computer vision task such as face recognition. In addition, due to the huge size of 3D MR image, training a deep DenseNet on the whole brain image not only requires high computation and memory costs but also inefficient to identify the abnormal changes relevant to AD. In this paper, we propose a classification framework by combination of multiple cluster DenseNets for classifications of AD vs. NC and MCI vs. NC, as shown in Fig.1. It consists of four main steps: 1) image acquisition and processing; 2) patch extraction and clustering; 3) construction of multiple cluster DenseNets; 4) final ensemble classification. These steps will be presented in the following subsections.

2.1. The image set and processing

The MR brain images used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of the ADNI was to test whether serial magnetic resonance imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as reduce the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California, San Francisco. ADNI was the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. The study subjects were recruited from over 50 sites across the U.S. and Canada and gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by each participating sites Institutional Review Board (IRB).

In this work, we use the T1-weighted MR imaging data from the baseline visits of 831 participants including 199 AD, 403 mild cognitive impairments (MCI) and 229 normal controls (NC) for evaluation. The details of the subjects used are shown in Table 1. In ADNI, the T1-weighted MR images are acquired sagittally using the volumetric 3D MPRAGE with $1.25 \times 1.25\,mm^2$ in-plane spatial resolution and 1.2 mm thick sagittal slices. Most of these images were obtained with 1.5T scanners, while a few were acquired using 3 T scanners. Detailed information about MR acquisition procedures is available at the ADNI website. The MR images are processed as follows. Specifically, all MR images are skull-stripped and cerebellum-removed after a correction of intensity inhomogeneity using a nonparametric nonuniform intensity normalization (N3) algorithm (Sled et al., 1998; Wang et al., 2011). The linear registration is performed to align the MR images to a template with FSL (FMRIB Software Library) 5.0 from https://fsl.fmrib.ox.ac.uk/.

2.2. Patch extraction and clustering

The goal of this step is to partition the whole MR brain image into a number of local regions and extract 3D patches from each region for clustering. To facilitate training the DenseNet and extraction of local features, we propose to uniformly partition the MR image into $3 \times 3 \times 3$ local regions of the same size. Since no rigid registration is performed on the images, there are some spatial shifts and deformations in the image regions which may result in different image spatial structure on the same region. In addition, the dementia of different subjects may have the abnormal changes in different brain regions. To alleviate these problems, we extract a number of 3D image patches of fixed size $32 \times 32 \times 32$ at a step of 2 voxels from each local region. Thus, there are a large number of 3D patches which are heterogenous with different spatial information. It is necessary to find the groups of patches with similar spatial structure to capture the characteristics of local regions. To achieve this, we apply K-means clustering method (Hartigan and Wong, 1979) to group the patches from each local region into K clusters. Since the length of vector by concatenating the $32 \times 32 \times 32$

Table 1
Demographic characteristics of the studied subjects from ADNI database (the values are denoted as mean ± standard deviation).

Diagnosis	Number	Age	Gender (M/F)	MMSE	Education	CDR
AD	199	75.43 ± 6.0	101/98	23.45 ± 2.1	14.66 ± 3.2	0.8 ± 0.25
MCI	403	74.9 ± 7.2	260/143	27.18 ± 1.7	15.75 ± 2.9	0.5 ± 0.03
NC	229	75.93 ± 5.0	119/110	28.93 ± 1.1	15.83 ± 3.2	0 ± 0

MMSE: the Mini-Mental State Examination; CDR: the Clinical Dementia Rating.

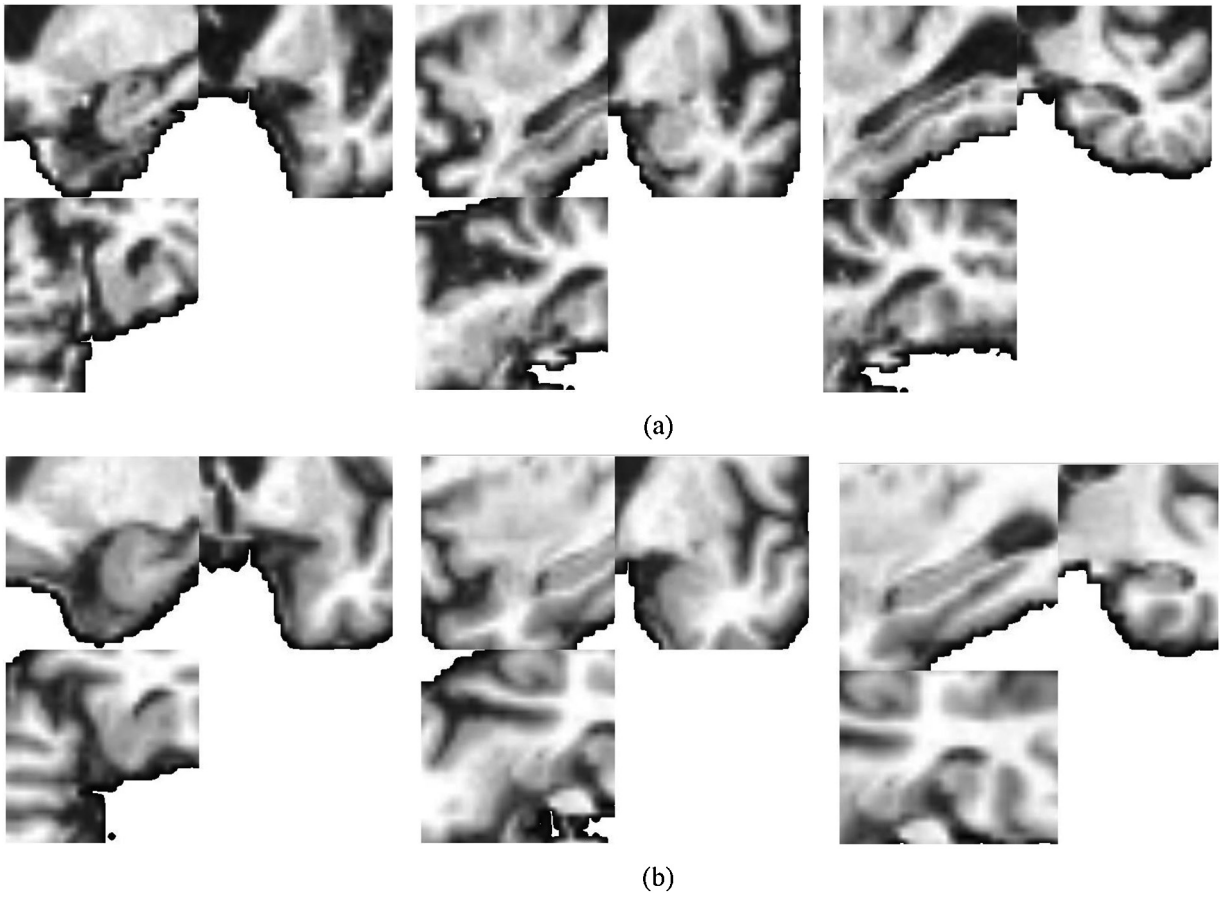

Fig. 2. The patches extracted at three different positions in the hippocampus region for (a) AD and (b) NC subject.

patch is very high, the computation efficiency is low for clustering. Principal component analysis (PCA) is used to reduce the dimension from 32,768 to 2000 for each patch. Then, the patches with the reduced feature dimension from the same region are grouped by K-mean clustering into many clusters with each cluster having similar image structure.

The number of clusters K has an important effect on clustering of different patches and the following feature extraction and classification. The variances of patches are usually caused by two main factors. First, the patches extracted from different positions may have different image structures resulting high variances. Second, the anatomical changes by disease cause variance of patches even from the same position. Fig. 2 (a) and (b) shows the patches extracted at three different positions in the hippocampus region for an AD and a NC subject, respectively. We can see that the variance of patches extracted at different positions is much larger than that of patches extracted at same position for different classes of subjects. It is better to capture the position variance using different clusters while the anatomical changes by disease still remain in the intra-clusters. If K is set too large, the clusters will be small and the patches within the clusters will have high similarity, which resulting in the small class discriminability for the patches from one cluster. In addition, large K requires more computation cost for clustering. On the contrary, if K is set too small, the clusters will be large consisting of many patches with large variance. This variance may be caused by different positions which reduces class discriminability of the patches within cluster. In our experiment, we extract the patches from 96 different positions and there is high similarity on the patches from the neighboring 9 positions. To balance this trade-off, K is set to about 10 in this work. In the following Section 3.4, we will test the proposed method with different numbers of clusters K. The results show that it can achieve optimal results by setting K to 10.

2.3. Multi-cluster dense convolutional networks

After patch extraction, features are extracted to capture the patch-level discriminative patterns of MR images. As mentioned above, the patches in one cluster have the similar spatial structure information in the close image positions. In addition, the patches of different clusters may have different discriminative power for AD classification. Convolutional neural networks (CNNs), which alternatively stacks several convolutional and pooling layers followed by fully connected and softmax layers, have been widely inves-

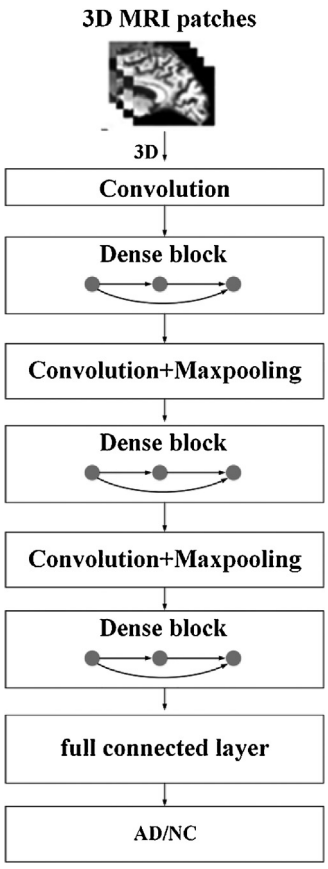

Fig. 3. The structure of proposed DenseNet with three dense blocks for each brain region.

tigated to learn features for image classification. To enhance the representation power, CNNs become increasingly deep so that the input information passes through many layers to reach the end of network. The low-level features will lose some important information when they are transformed to the high-level layer. To address this problem, the Dense Convolutional Network (DenseNet) was proposed to connect each layer to every other layer in a feed-forward fashion, which increases direct connections between the low and high level layers (Huang et al., 2016). In the DenseNet, the layers from different levels are connected densely to improve information flow between layers. Compared to the traditional deep CNNs, the DenseNets have several compelling advantages (Huang et al., 2016). First, they can alleviate the vanishing-gradient problem since there is a direct connection from the low to high level layers. Second, feature propagation is strengthened to encourage the reuse of some features especially the low-level features. Third, they can substantially reduce the number of network parameters. Thus, DenseNet can achieve better performance than the traditional deep CNN for image classification.

Motivated by the success of DenseNets, we propose to construct a 3D DenseNet with the 3D image patches from each cluster to learn the patch-level representations of local image regions. Thus, multiple cluster DenseNets are constructed to capture the abnormal changes of the local brain region caused by dementia and aggregated for representation of the whole brain. The multiple cluster DenseNets share the same deep network structure as shown in Fig. 3. The detailed parameters of the DenseNet are shown in Table. 2. Since different clusters capture different spatial image structures of local regions, a DenseNet is trained with all the patches from one cluster, and the patches' label is same as their subject label. Thus, multiple DenseNets have different network

Table 2
The architecture of the DensNet.

Layers	Output Size	Filter size, stride, number
Input Layer	$32 \times 32 \times 32$	–
Convolution	$32 \times 32 \times 32$	$3 \times 3 \times 3$, 1, 16, conv
Dense Block (1)	$32 \times 32 \times 32$	$3 \times 3 \times 3$, 1, 12, conv
Transition Layer (1)	$32 \times 32 \times 32$	$3 \times 3 \times 3$, 1, 28, conv
	$16 \times 16 \times 16$	$2 \times 2 \times 2$, maxpooling
Dense Block (2)	$16 \times 16 \times 16$	$3 \times 3 \times 3$, 1, 12, conv
Transition Layer (2)	$16 \times 16 \times 16$	$3 \times 3 \times 3$, 1, 40, conv
	$8 \times 8 \times 8$	$2 \times 2 \times 2$, maxpooling
Dense Block (3)	$8 \times 8 \times 8$	$3 \times 3 \times 3$, 1, 12, conv
Global Average Pooling	52	–
Full-connected layer	10	–
Softmax Layer	2	–

weights to learn the specific discriminative features of different clusters. The DenseNet model contains three dense blocks, and each block consists of three layers: one convolutional layer, one batch normalization layer, and one activation layer (Relu). For convolutional layers, each side of the inputs is zero-padded by one pixel to keep the feature-map size fixed, and all layers have the same feature-map sizes and are directly connected with each other. The layers between two dense blocks are transition layers to reduce the feature-map size, which consists of four layers: one convolutional layer, one dropout layer, one maxpooling layer and one batch normalization layer. The input and output of dense blocks are concatenated as the input of transition layers. After each convolutional layer, output size is same as input size, and the maxpooling layers are used to down-sample the input feature map along the spatial dimensions. The fully connected layer consists of a number of output neurons. Each neuron outputs the learned linear combination of all the inputs from the previous layer and passed through a nonlinearity. All layers receive the direct supervision from the loss function through the shortcut connections. Each DenseNet is optimized individually for the classification task by a softmax layer to output the class probabilistic score. Thus, the subject label information is used through back-propagation for learning the most relevant features and updating the weights of DenseNet model. The output of last fully connected layer in each cluster DenseNet is considered as the learned patch-level representation features for the cluster. After training of cluster DenseNets, we select the most discriminative DenseNets with high classification accuracies on the validation sets, from which the learned patch-level features will be combined for the representation of local brain regions.

2.4. The final ensemble classification

For final image classification, we gradually combine the multiple DenseNets trained for different clusters to generate the final ensemble classification of MR brain images. The patch-level features from selected clusters are aggregated to cluster and regional representations which are finally ensembled for global image level classification. This stage is an ensemble stage and it contains two parts. One part is the combination of cluster outputs for representation of each local region. Another one is the ensemble of outputs from different local image regions for final global image classification.

First, for each image region, different clusters may have different discriminative power for AD classification. Thus, we choose several clusters with higher classification accuracy on the validation data for ensemble. In addition, different subject may have different numbers of patches assigned for each cluster. To facilitate further processing, we need to generate a feature vector of fixed length with the DenseNet for representation of each image region. Let X_{ijk} denote the output features of patch k in the cluster j of subject i. For

cluster j, assume that the subject i have K patches, the aggregated features of subject i for the cluster are calculated as:

$$X_{ij} = \sum_{k=1}^{k=K} X_{ijk}/K \quad (1)$$

where X_{ij} is the output features of cluster j for subject i, and X_{ijk} is the output of fully connected (FC) layer in the DenseNet for patch k. If there are no patches from subject i in cluster j, we use zeros to replace the output of FC layer. In this way, the patch-level representation features are aggregated into a cluster level representation.

There are multiple clusters grouped for each brain region with each cluster capturing the specific pattern. After aggregating the patch-level features for each selected cluster, the features of multiple discriminative clusters are further integrated for representation of the local image region. To achieve this, we concatenate the output features of the selected clusters as one vector:

$$X_i = H[X_{i1}, \ldots, X_{ij}] \quad (2)$$

where X_i is the representation feature vector for subject i and X_{i1} is the feature vector for the first chosen cluster, and H is the concatenating function. After generating the features for each image region, we append a fully-connected and a softmax layers with the task of AD classification. The fully-connected and softmax layers are trained with back-propagation for learning the most relevant features to generate the class prediction scores for the region. Thus, the cluster-level features are combined for regional discrimination.

Second, we ensemble the predict scores of different image regions for final classification of whole brain image. We concatenate the predict scores of image regions into a feature vector. And a two-layer network with a full connected layer and a softmax classification layer is finely trained to make the final classification for each subject. Our proposed method can gradually aggregate the patch level information from cluster level to region level and finally generate the global image level classification. In our deep learning model, the multiple cluster DenseNets extract discriminative image features, while the upper fully connected and softmax layers facilitate the task-specific classification. Thus, training of the proposed method consists of pre-training individual DenseNets, and the task-specific fine-tuning for ensemble classification. Initially, a 3D DenseNets is pre-trained for each cluster by directly mapping the outputs of the fully connected layer to the probabilistic scores of all class labels with a softmax function. Then, the patch-level features are integrated into regional features to train the upper fully connected and softmax prediction layers for the classification task.

In summary, our proposed classification method is a deep learning based model, which can gradually learn the features of MR brain images from the local patches to global image level. Specially, we learn the patch-level features using multiple cluster DenseNets to capture the local spatial structure information, which are further aggregated to the regional representation and ensembled for final image level classification. These multi-level feature learnings are incorporated into the classification task learning process. Compared to the existing methods for classification of MR brain images, our proposed method has the following advantages. First, no segmentation and rigid registration are required in image processing before feature extraction, which can simplify the diagnosis procedure and save the computation costs. Second, it is a data driven method by jointly learning the features and classification task without requiring the prior knowledge of domain experts. Third, instead of training one DenseNet on the whole brain image, the proposed method trained multiple cluster DenseNets to learn the features of different local image regions with the sampled image patches. This can address the problem of small image set on training a deep learning model.

3. Experimental results

In this section, we will first introduce the image datasets and implementation of our proposed method. Then, we will present the extensive experiments to test the proposed method on classifications of AD vs. NC and MCI vs.NC. We will further compare our proposed method with other methods reported in the literature and give the discussion.

3.1. Datasets and implementation

The proposed classification method is tested on the T1-weighted MR brain images from ADNI database. The MR images are taken from the baseline visits of 831 participants including 199 AD, 403 MCI and 229 NC for evaluation. The proposed method is tested on classifications of AD vs. NC and MCI vs. NC. The image processing is conducted as illustrated in Section 2. The size of MRI image after processing is $256 \times 256 \times 256$ voxels. We down-sample the images by 2 and remove the voxels whose intensity values are zeros. The brain images of size $98 \times 78 \times 76$ voxels are obtained for test and this image is uniformly partitioned into $3 \times 3 \times 3$ regions. A number of 3D image patches are extracted from each region and are grouped into many clusters. And a deep 3D DenseNet is trained for each cluster. After training of multiple 3D DenseNets, the discriminative cluster DenseNets with high classification accuracies on the validation sets are considered for further processing. The proposed method is implemented with python 2.7.9 and the DenseNets are constructed with the Keras library in the framework of Tensor flow. For training the DenseNets, the initial weights for whole network is uniform, which is default in Keras. Adam optimizer is adopted with a low learning rate of $1 \times 10\text{-}4$, The networks are stable after iteration of 70 epochs. The batch size is set to 128. *Relu* activation is used for each neuron of DenseNets. All the experiments are conducted in the environment of Ubuntu14.04-x64/ GPU of NVIDIA GeForce GTX 1080 Ti. The L2 regulation and dropout are also used to address the overfitting problem.

To evaluate the classification performance, we use a 5-fold cross-validation strategy to train and test the proposed method to reduce the effects of random factors. Each time, one fold of data set is used for testing, while the other folds were used for training. The training set is further split into training and validation parts. The validation part is used for finetuning the iteration numbers of the training process to obtain the model weights with the optimized performance. In our experiments, four classification performance measures are used for evaluation including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic curve (ROC) and area under ROC curve (AUC). ACC is computed as the proportion of correctly classified subjects among the whole population. SEN is computed as the proportion of correctly classified positive samples (AD/MCI subjects) among the total number of positive samples. SPE is the proportion of correctly classified negative samples (NC subjects) among the total number of negative samples.

3.2. Test the effectiveness of different number of clusters K

As stated in Section 2, the number of clusters K has important effect on the following feature extraction and classification. Small K cannot partition the patches from different locations very well while large K will generate small clusters with no enough patches from different subjects for training the DenseNet. Thus, we test the proposed method by adjusting the number of clusters K by 5, 10, 15 to compare their performances. First, we test the classification accuracy of each individual cluster with the DenseNets for different number of clusters. Fig. 4 shows the comparison of classification accuracy of the *K (K=5, 10, 15)* clusters with the DenseNets for

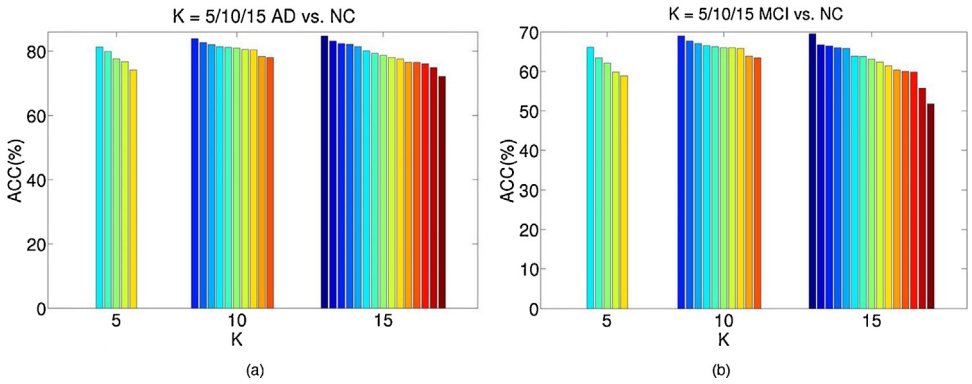

Fig. 4. Comparison of classification accuracies of K (K = 5, 10, 15) clusters with the DenseNets for classifications of (a) AD vs. NC and (b) MCI vs. NC, where each bar denotes the classification accuracy of one cluster.

Table 3
Comparison of classification accuracies with different *K* for classifications of AD vs. NC, and MCI vs. NC.

K	AD vs. NC (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
5	88.1	88.0	88.2	92.3	70.6	78.2	57.2	72.9
10	89.5	87.9	90.8	92.4	73.8	86.6	51.5	77.5
15	89.7	86.4	92.6	93.7	74.0	86.6	52.0	73.7

Table 4
Comparison of the classification performances with three level features for classifications of ad vs. nc and mci vs. nc.

Methods	AD vs. NC (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Cluster level	84.0	77.1	90.0	90.4	69.0	86.1	39.6	71.9
Region level	85.2	88.2	81.9	92.3	70.5	89.7	35.4	72.0
Image level	89.5	87.9	90.8	92.4	73.8	86.6	51.5	77.5

classifications of AD vs. NC and MCI vs. NC. We can see that the classification performance is poor for $K = 5$ and it is improved by increasing K to 10. When K is increased to 15, the classification performances of some clusters are significantly reduced. The average classification performance of K clusters is optimal with $K = 10$.

In addition, we also compare the final ensemble classification performances of the proposed method by setting different number of clusters as shown in Table 3. From these results, we can see that the classification accuracies are improved about 1.4% and 3.2% by increasing K from 5 to 10 for classifications of AD vs. NC and MCI vs. NC, respectively, but they cannot be further improved much by further increasing K from 10 to 15. As for the computation cost, the proposed method includes offline training and online testing processes. The offline training process includes the K-means clustering and DenseNets training. Large K will increase the computation of clustering. Since we need to train a DenseNets for each cluster, more computation is required to train the DenseNets for large K. The K-means clustering is implemented with MATLAB software while the DenseNets is implemented with Python in our experiments. The offline training process takes 2.6, 3.3, 4.6 h for $K = 5$, 10 and 15, respectively. The online test for one subject takes about 0.42, 0.84, 1.26 s on average for $K = 5$, 10 and 15, respectively. Based on these results, K is set to 10 in the experiments.

3.3. Test the effectiveness of representation features at different levels

In this section, we perform the experiments to test the effectiveness of the representation features at different levels. For fair comparison, the only difference is the features used for classification and the following classifier models are same. We compare the classification performances using the representation features learned from the cluster level, region level and final image level. For the cluster level, we select the cluster DenseNet with the best classification performance for comparison, denoted as "Cluster level". For the region level, we select the image region with the best classification performance for comparison, denoted as "Region level". The "Image level" represents the final classification performance of the whole brain image by the proposed method. Table 4 demonstrates the comparison of classification performances at three different levels for AD vs. NC and MCI vs. NC. Furthermore, the ROC curves for classifications of AD vs. NC and MCI vs. NC are shown in Fig.5 (a) and (b), respectively. From these results, we can see that the representation features of region level can improve the classification performances about 1.2% by aggregating the representation features of cluster level for classification of AD vs. NC. By ensemble of region level representations, the final image-level classification accuracy is improved about 4.3% and 3.3% compared to the region level classification for classifications of AD vs. NC and MCI vs. NC, respectively.

3.4. Comparison with other methods

In this section, we conduct the experiments to compare the proposed method with other methods. For fair comparison, the experiments are performed on the same training and test data set. The first experiment is to compare the proposed method between the DenseNet with other deep learning models. We conduct the experiments by replacing the DenseNet with other well-known deep learning models such as LeNet-5 (Lécun et al., 1998) and ResNet (He et al., 2016). All other steps such as cluster selection and ensemble classification are same as our proposed method. The deep LeNet-5 consists of three convolutional layers with each followed by a subsampling layer, one fully connected layer and an output layer of Gaussian connection (Lécun et al., 1998). The ResNet is a residual learning framework to ease the training of deeper networks than those used previously (He et al., 2016). The ResNet achieved the best classification performance in the 2015 ImageNet competition. The layers are reformulated as learning residual functions with reference to the layer inputs. In our experiments, we download their source codes and implement these methods with our best efforts by modifying the 2D convolutions with the 3D convolutions. The classification results for AD vs. NC and MCI vs. NC are compared in Table 5. We can see that DenseNet performs better than the LeNet-5 and ResNet because it can make full use of multi-layer rich features.

The second experiment is to compare our proposed method to some other methods that are also based on T1-weighted struc-

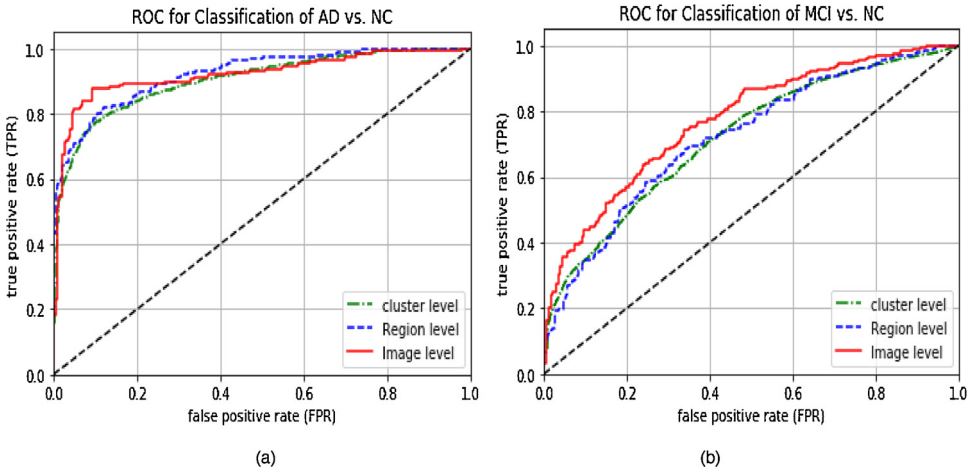

Fig. 5. Comparison of ROC curves on the three level features for classifications of (a) AD vs. NC and (b) MCI vs. NC.

Table 5
Comparison of the classification performances between DenseNet and other deep learning methods.

Methods	AD vs. NC (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
LeNet-5	87.6	82.5	91.7	90.7	71.1	86.1	44.6	72.8
ResNet	88.3	86.0	90.4	90.4	72.9	78.9	62.4	71.5
DenseNet	89.5	87.9	90.8	92.4	73.8	86.6	51.5	77.5

Table 6
Comparison of the classification performances for different feature extractions.

Methods	AD vs. NC (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
ROI feature	81.7	67.1	94.3	88.3	66.8	78.9	41.3	70.2
Voxel-wise feature	86.4	79.3	92.5	91.9	71.5	82.8	51.5	75.6
One DenseNet	85.5	83.4	87.3	88.7	70.3	77.9	56.8	76.0
Our method	89.5	87.9	90.8	92.4	73.8	86.6	51.5	77.5

tural MRI data of ADNI. First, we compare our proposed method to three state-of-art methods, which compute the GM volumes of 93 ROIs for representation as (Zhang et al., 2011), compute the voxel-wise grey matter density maps for features as (Liu et al., 2012), and build a DenseNets on the whole brain image. For fair comparison, the same image preprocessing and classifier, i.e., softmax classifier, are used for all compared methods with only difference on feature extraction in the experiments. For extraction of ROI and voxel-wise features, the rigid registration and tissue segmentation are required. In our experiments, FAST in the FSL package (Zhang et al., 2001) is used to segment the cerebellum removed MRIs into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The rigid registration tool of HAMMER (Shen and Davatzikos, 2002) is used for the nonlinear MR image alignment and mapping the image into 93 manually labeled ROIs (Kabani et al., 1998). For each labeled MR image, we compute the normalized volumes of GM tissue in all ROI regions as the features. For extraction of voxel-wise features, the same registration and tissue segmentation methods as ROI based methods are used for fair comparison. After the image warping by HAMMER (Shen and Davatzikos, 2002), the tissue volume within any size of region is preserved and the warped mass-preserving tissue volumes reflect the spatial distribution of tissues in an original brain called as the tissue density maps. The spatially normalized GM density maps are used as voxel-wise features for classification. In addition, one direct deep network model is to build a DenseNet on the whole brain image. Since the whole brain image size is of $256 \times 256 \times 256$ voxels, it is too large to train the DenseNet with our GPU GTX 1080 Ti because of the memory limit. Thus, we remove the voxels of zeros values and down-sample the whole brain images by 2 to obtain the brain images of $98 \times 78 \times 76$ voxels for training the DenseNet. This can not only address the problem of memory limit but also alleviate the overfitting problem caused by training a large number of deep network parameters with a small number of training subjects. Table 6 shows the comparison of the classification performances with these different methods. From these results, we can see that the voxel-based method performs better than the ROI-based method and one DenseNet. Nevertheless, our proposed method achieves a very competitive classification accuracy of 89.5% for classification of AD vs. NC, which is higher than the voxel-based method (86.4%). For classification of MCI vs. NC, our method also achieves the accuracy of 73.8%, higher than the voxel-based method (71.5%). It is worthy to note that the evaluation here is on the feature extraction, not the design of classifier, so the results may be different from those reported in literature (Zhang et al., 2011; Liu et al., 2012).

3.5. Discussion

Instead of extracting the handcrafted features from MR images, our proposed method constructs multiple cluster DenseNets to gradually learn the multi-level features for global image classification. Each DenseNet captures the patch-level image features which are combined to generate higher-level features to achieve more robust classification. There are no tissue and ROI segmentations and no rigid registration required in processing the brain images so that the computation cost is reduced. However, there are still some limitations in the proposed method. First, the parameters of the DenseNet model, such as the number and types of layers, may not be optimally determined. This problem can be alleviated by optimizing the parameters using the validation data set. Second, it is not easy to visualize the learned features by the proposed method for interpretation of the brain regions relevant to neurodegenerative disease (i.e., AD or MCI) for the clinical application. The learned features have no sufficient clinical information to find the related ROIs for clinical understanding of brain abnormalities. To facilitate the interpretation of diseases, we select the top clusters with high classification accuracies because these cluster have the large discriminative power. For each subject, the patches included in the selected discriminative clusters are considered as the relevant regions. Thus, we can obtain a relevance map by compiling the masks generated by all the selected image patches for each subject.

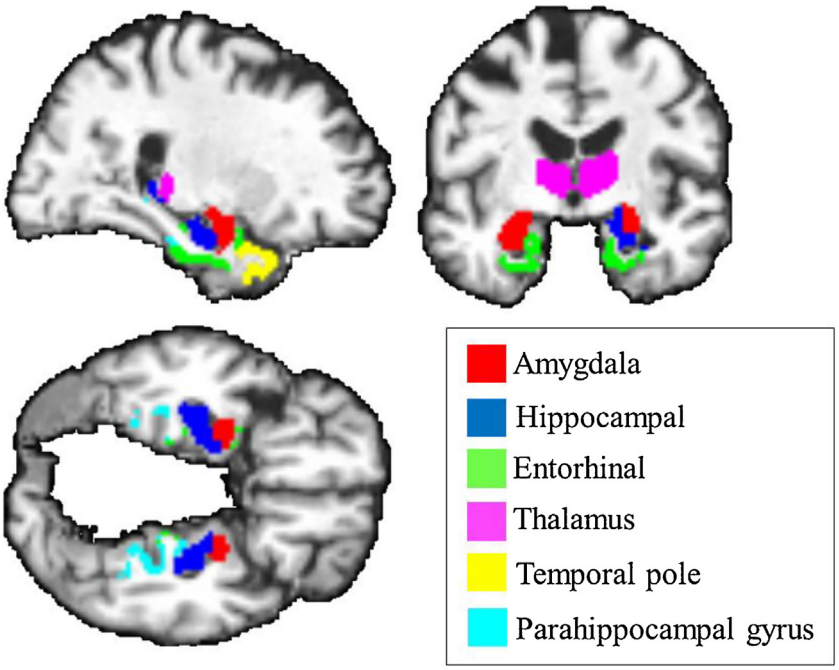

Fig. 6. The top relevant regions of interests (ROIs) in brain covered by the discriminative patches in AD diagnosis.

This map shows the importance of ROIs in prediction of disease status. To facilitate the interpretation, we map the relevance map to a template with 93 manually labeled ROIs for reference (Kabani et al., 1998). The more patches are mapped to the ROI for different subjects, the higher probability the ROI have to be relevant to disease. For each ROI, we sum the proportion of the voxels covered by the relevance maps from different subjects. Fig. 6 shows the top 6 ROIs of brain with the high relevance to AD diagnosis. These areas seem to be consistent with the ones that are mostly affected by AD, mainly in amygdala, hippocampus, entorhinal, thalamus, temporal, parahippocampal gyrus (Zhang et al., 2011; Liu et al., 2015a).

As for the computation cost, the proposed classification algorithm consists of both the offline training and online testing stages. In the training stage, the computation includes the K-means clustering and DenseNets training, which take 0.5 h and 2.8 h in our experiments, respectively. In the testing stage, computation cost is 0.84 s to test the proposed algorithm for a given image. The experiments are all conducted on PC with GPU NVIDIA GeForce GTX1080 Ti of 12GB memory.

4. Conclusion

This paper has proposed a classification method based on combination of multiple cluster DenseNets for AD and MCI diagnosis using MR brain images. The whole brain image is partitioned into a number of local regions and a number of 3D patches are extracted from each image region. K-means clustering is used to group patches with similar spatial structure into several clusters and a DenseNet is built and trained for each cluster to extract the patch-level features. The features learned by multiple cluster DenseNets are gradually aggregated for image classification. The proposed method is a learning based method to jointly learn the feature representation and classification without prior domain knowledge. No rigid registration and segmentation is required in the image processing. Experimental results and comparison demonstrate that the proposed method can solve the small training set problems and achieve higher accuracy to classify the AD from NC and MCI from NC using structural MRI scans. Comparison with the existing methods shows the promising classification performances for AD diagnosis. In the future work, the proposed method can be extended to other brain imaging modality such as PET image and extended for multimodal brain image analysis.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (NSFC) under grants (No. 6181101049, 61773263, 61375112), The National Key Research and Development Program of China (No.2016YFC0100903) and SMC Excellent Young Faculty program of SJTU.

References

Adrien, Pa.G.M., arXiv:1502.02506 [cs.CV] 2015. Predicting Alzheimer's Disease: A Neuroimaging Study with 3d Convolutional Neural Networks.

Ben, A.O., Mizotin, M., Benois-Pineau, J., Allard, M., Catheline, G., Ben, A.C., 2015. Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. Comput. Med. Imaging Graph. 44, 13–25.

Cao, P., Liu, X., Yang, J., Zhao, D., Huang, M., Zhang, J., Zaiane, O., 2017. Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures. Comput. Biol. Med. 91.

Cheng, D., Wang, Y., 2017. Classification of MR brain images by combination of multi-CNNs for AD diagnosis. In: International Conference on Digital Image Processing.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a K-Means clustering algorithm. Appl. Stat. 28 (1), 100–108.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition.

Herrup, K., 2011. Commentary on R̈ecommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.Ädressing the challenge of Alzheimer's disease in the 21st century. Alzheimers Dementia J. Alzheimers Assoc. 7 (3), 335.

Hosseini-Asl, E., Keynton, R., El-Baz, A., 2016. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. In: 2016 IEEE International Conference on Image Processing (Icip)., pp. 126–130.

Huang, G., Liu, Z., Weinberger, K.Q., 2016. Densely Connected Convolutional Networks. CVPR.

Jr, J.C., Albert, M.S., Knopman, D.S., Mckhann, G.M., Sperling, R.A., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dementia J. Alzheimers Assoc. 7 (3), 257.

Kabani, N., MacDonald, D., Holmes, C.J., Evans, A., 1998. A 3D atlas of the human brain. Neuroimage 7 (4), S717.

Le, A., Adeli, E., Liu, M., Zhang, J., Lee, S.W., Shen, D., 2017. A hierarchical feature and sample selection framework and its application for Alzheimer's disease diagnosis. Sci. Rep. 7, 45269.
Lécun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Liu, M., Zhang, D., Shen, D., 2012. Ensemble sparse classification of Alzheimer's disease. Neuroimage 60 (2), 1106–1116.
Liu, F., Suk, H.I., Wee, C.Y., Chen, H., Shen, D., 2013. High-order graph matching based feature selection for Alzheimer's disease identification. Springer, Berlin Heidelberg.
Liu, F., Wee, C.Y., Chen, H., Shen, D., 2014. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. Neuroimage 84, 466–475.
Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015a. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans. Biomed. Eng. 62 (4), 1132–1140.
Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015b. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans. Biomed. Eng. 62 (4), 1132.
Liu, M., Zhang, J., Nie, D., Yap, P.T., Shen, D., 2018. Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. IEEE J. Biomed. Health Inform. 99, 1-1.
Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21 (11), 1421.
Shen, D., Wu, G., Suk, H.I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87.
Suk, H.I., Lee, S.W., Shen, D., 2015. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct. Funct. 220 (2), 841–859.
Suk, H.I., Lee, S.W., Shen, D., 2017. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med. Image Anal. 37, 101.
Wang, Y., Nie, J., Yap, P.T., Shi, F., Guo, L., Shen, D., 2011. Robust deformable-surface-based skull-stripping for large-scale studies. Med. Image Comput. Comput. Assist. Interv. 14 (3), 635–642.
Ye, T., Zu, C., Jie, B., Shen, D., Zhang, D., 2016. Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease. Brain Imaging Behav. 10 (3), 739.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20 (1), 45–57.
Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55 (3), 856.
Zhang, J., Gao, Y., Gao, Y., Munsell, B., Shen, D., 2016. Detecting anatomical landmarks for fast alzheimer's disease diagnosis. IEEE Trans. Med. Imaging 35 (12), 2524–2533.
Zhang, J., Liu, M., An, L., Gao, Y., Shen, D., 2017. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. IEEE J. Biomed. Health Inform. 21 (6), 1607.
Zhu, X., Yao, J., Zhu, F., Huang, J., 2017. WSISA: making survival prediction from whole slide histopathological images. In: IEEE Conference on Computer Vision and Pattern Recognition.